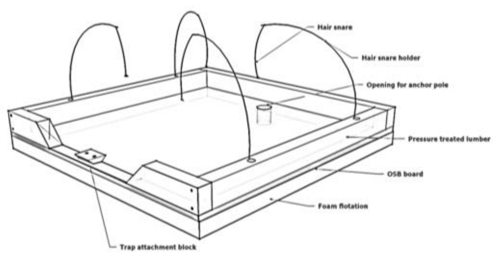
Single Season, Multi-Method Occupancy Case Study – Estimating Occupancy of Nutria in Maryland.

Project Description and Context

In this module, we'll fit occupancy models using covariates to evaluate the effectiveness of three detection methods: hair snares, presence of scat, and trail cameras. The species of interest is Nutria (*Myocaster coypus*), a semi-aquatic rodent native to South America which was introduced into Maryland in 1943. The original study objectives were to: 1) determine if platform placement on land or water influenced nutria visitation rates, 2) determine if the presence of hair snares influenced visitation rates, and 3) determine method-specific detection probabilities.



This study was conducted from July through October 2012 on the Wicomico River, Maryland. Platforms were built for the collection of hair and scat, with cameras mounted nearby to observe platform activity. 40 platforms were checked every 2-4 days from 4 September until 26 October (15 checks). See Pepper et. al. for more details.

The data are contained in a spreadsheet, *NutriaData.csv*. Column labels indicate the survey (check) and method (x.y : x=survey, y=method) and the covariate to indicate whether the platform was placed on land or water (0=land, 1=water). Note here which method was which number in the spreadsheet (1,2, or 3).

References:

Pepper, M.A., Herrmann, V., Hines, J.E., Nichols, J.D. and Kendrot, S.R. (in prep).

Exercise Objectives

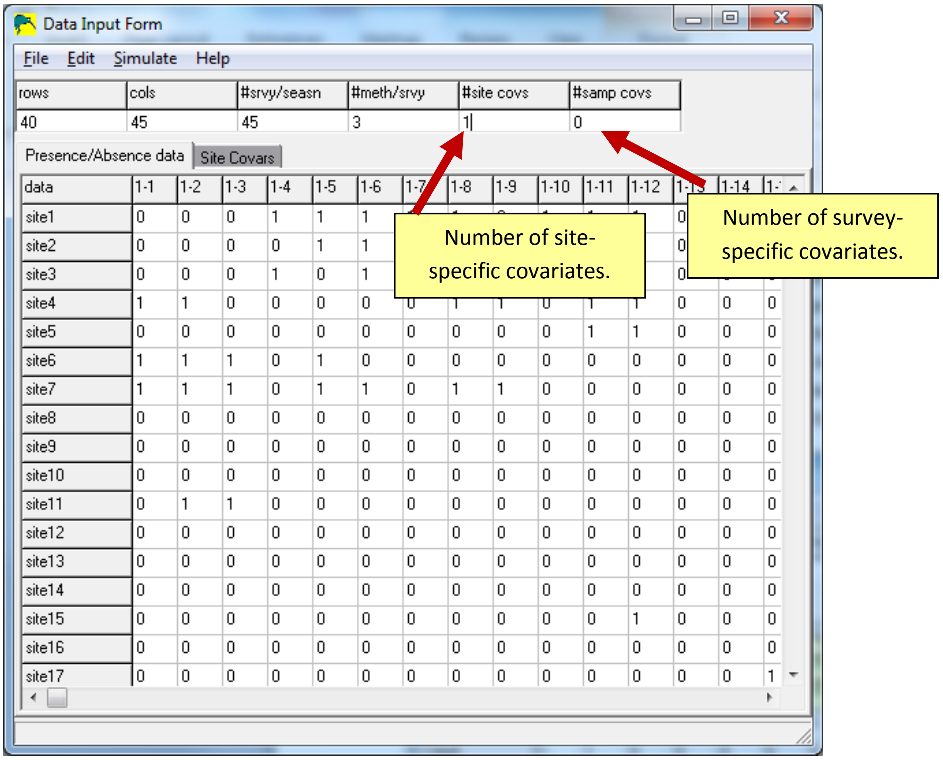
- Learn how to create and run multi-method occupancy models where occupancy and/or detection is a function of site and/or survey specific covariates
- Learn to import and analyze multi-method data
- Continue to increase comfort level and familiarity with all aspects of analysis in PRESENCE from data exploration to model selection, data filter assignment and interpretation of results

Single Season, Multi-Method Occupancy Case Study – Estimating Occupancy of Nutria in Maryland.

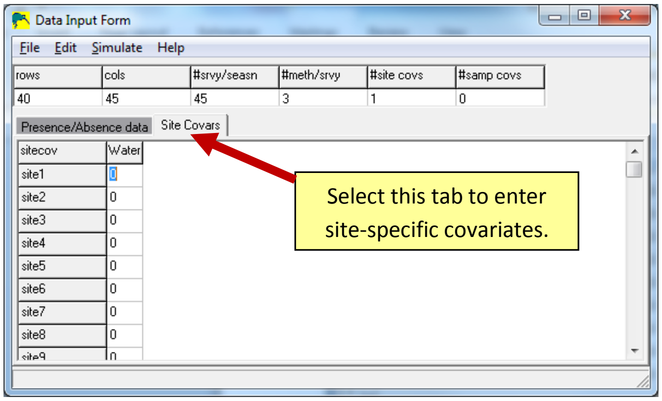
Presence spreadsheet data file: ***NutriaData.csv***

INSTRUCTIONS

Step 1 – Data Import: Begin PRESENCE, start a new project and open the data input form (click "Data Input Form" button). Copy and paste the detection-nondetection data from the spreadsheet into PRESENCE (highlighting on the values and using Edit>Paste>Paste values), and change the number of methods per survey to 3. *(The methods used were: scat, camera, and hair-snare, which correspond to the 3 columns for each survey.)* Next, we have 1 site-specific covariate (whether a site was on land or in water), so change the number of site-specific covariates to 1. As the site-specific covariates must remain constant for the entire season, each one only requires 1 column hence all site-specific covariates are entered on a single sheet. The sampling occasion covariates can vary with each survey; hence the covariate will have the same dimensions as the detection-nondetection data. Therefore, 1 sheet is required by each sampling covariate.



Single Season, Multi-Method Occupancy Case Study – Estimating Occupancy of Nutria in Maryland.



To input the site-specific covariate, in your spreadsheet program select the last column (including the column label, "Water") (i.e., cells AU1:AU41), copy the selected cells, return to PRESENCE's data input form and select the tab labelled "**Site Covar"**. Click on the top-left grid cell and make sure the cursor is in that cell, then from the edit menu, select **Edit>Paste>Paste w/covnames**. This paste option will not only paste the covariate values into the grid cells, but will also paste in the covariate names.

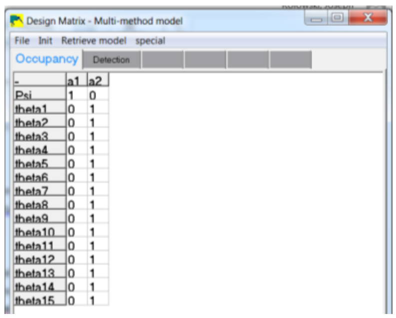
Once you have entered the data and covariates, save the data file using an appropriate name (here I have used 'NutriaExample') and location, then close the data input form which will return you to the **Enter Specifications for PRESENCE Analysis** window. Click 'OK' to create a project folder which will contain the data and all models. After a couple of seconds a blank results browser should appear (i.e., an empty table containing only column headings). Importantly, if you do not see the results browser, you have not successfully set up your project file.

Step 2 – Data Exploration:

Bring back the Data window (**View>Data**) and examine which cells contain missing values for the Presence/Absence data.

Note that during the last 3 check periods, some sites were not sampled. It is important to distinguish these cases by denoting them with the missing value character, '-'.

Step 3 – Running a simple model: First, let's fit a simple model where the probability of occupancy is the same for all plots and the probability of detection is the same in all surveys. Click **Run>Analysis:single-season>Multi-method.** In our earlier notation we could call this model **psi(.)theta(.)p(.)**. The design matrix for occupancy will be a 16 X 2 matrix, with a '1' in the first row, first column, and 1's in the 2nd column of all other rows. You'll need to delete the remaining columns. This means global occupancy will be estimated using one parameter (a1), and all local occupancy estimates will be estimated using another parameter (a2). Your matrix should look as shown below:



The Detection design matrix will be a 45 X 1 matrix, with 1's in all cells. This means detection probabilities for all surveys and all methods are the same (estimated with parameter, b1). Click the 'Detection' tab and delete the 2nd and 3rd columns (left-click in col 2, right-click, then select 'Del Cols'). Change the 1st column to all 1's (**Init>Constant**). Go back to the previous window (don't close the design-matrix window), change the model name to 'psi(.)theta(.)p(.)' and click 'Ok to Run'. Click 'Yes' to add the results of the model to the results browser.

Step 4 – Analyze the influence of method on detection: Next we'll fit a model where the probability of detection varies among detection methods. We could call this model **psi(.),theta(.)p(m)**. Start the analysis by selecting **Run>Analysis:single-season>Multi-method**.

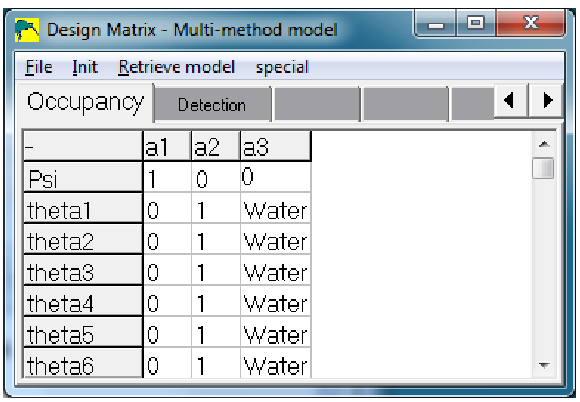
In the occupancy design-matrix, global occupancy is estimated using parameter, a1, and local occupancy for all surveys is estimated using parameter, a2 (all local occupancy estimates equal). This should look the same as in the first model you set up. In the detection design-matrix, detection probabilities are different for each method, but equal among surveys. So, there are 3 parameters to estimate: b1-detection prob. for method 1, b2-detection prob. for method 2, and b3-detection prob. for method 3. This should be the default setting when you open a new design matrix.

Once the design matrices are set up, return to the SNER window (without closing the design matrix window), rename the model then hit 'OK to Run' and confirm the results when prompted.

Examine the output for this model, paying attention to the values for probability of detection. Notice that the same 3 values for detection probability are repeated for each of the 15 surveys/checks. Which method is best/worst?

Step 5 – Analyze the influence of water on local occupancy (theta):

Next we'll fit a model where the probability of local occupancy will be different for platforms on land versus platforms in water. We'll assume local occupancy doesn't change from one survey to the next (constant) and detection probabilities are method-specific. We could call this model **psi(.)theta(water),p(m)**. We'll first fit the model, then work back through it explaining what we're actually doing. So start the analysis by selecting **Run>Analysis:single-season>multi-method**. Retrieve the last model you ran by selecting **Retrieve Model** from the design matrix window and selecting your previous model. In the design-matrix window, click the 'Occupancy' tab, then right-click and select 'Add col'. Click in the 3rd column, then click the '**Init>Indiv.Covariates>*Water**' menu. Finally, change the 'Water' to '0' in the 1st row. The design matrix window for occupancy will look like the following.



You can type the name of the covariate into each of the grid cells, but if you do so you must get the name 100% correct (and note that PRESENCE is case sensitive).

Once the design matrices are set up, return to the SNER window (without closing the design matrix window), rename the model (psi(.)theta(water)p(m)), then hit 'OK to Run' and confirm the results when prompted.

So what have we just done? As in the previous model, detection probabilities are method-specific. Global occupancy is the same for all sites, but local survey occupancy can be different, depending on whether the platform is on land or in water. To read the design matrices we move along the row, sum the terms produced by multiplying the values in the grid cell with the corresponding beta parameters for each column. So here we could write the set of equations:

Single Season, Multi-Method Occupancy Case Study – Estimating Occupancy of Nutria in Maryland.

$$\text{logit}(\mathit{psi}) = \mathbf{1} \text{ X } a_1 + \mathbf{0} \text{ X } a_2 + \mathbf{0} \text{ X } a_3 = a_1$$

$$\text{logit}(\theta_{l,site1}) = \mathbf{0} \text{ X } a_1 + \mathbf{1} \text{ X } a_2 + \mathbf{Water}_{\text{site1}} \text{ X } a_3 = a_2 \text{ [site on land]}$$

$$\text{logit}(\theta_{l,site2}) = \mathbf{0} \text{ X } a_1 + \mathbf{1} \text{ X } a_2 + \mathbf{Water}_{\text{site2}} \text{ X } a_3 = a_2 \text{ [site on land]}$$

: : :

$$\text{logit}(\theta_{l,site10}) = \mathbf{0} \text{ X } a_1 + \mathbf{1} \text{ X } a_2 + \mathbf{Water}_{\text{site10}} \text{ X } a_3 = a_2 + a_3 \text{ [site in water]}$$

: : :

For detection, the design-matrix translates to these equations:

$$\text{logit}(p_{1,method1}) = \mathbf{1} \text{ X } b_1 + \mathbf{0} \text{ X } b_2 + \mathbf{0} \text{ X } b_3 = b_1$$

$$\text{logit}(p_{1,method2}) = \mathbf{0} \text{ X } b_1 + \mathbf{1} \text{ X } b_2 + \mathbf{0} \text{ X } b_3 = b_2$$

$$\text{logit}(p_{1,method3}) = \mathbf{0} \text{ X } b_1 + \mathbf{0} \text{ X } b_2 + \mathbf{1} \text{ X } b_3 = b_3$$

: : :

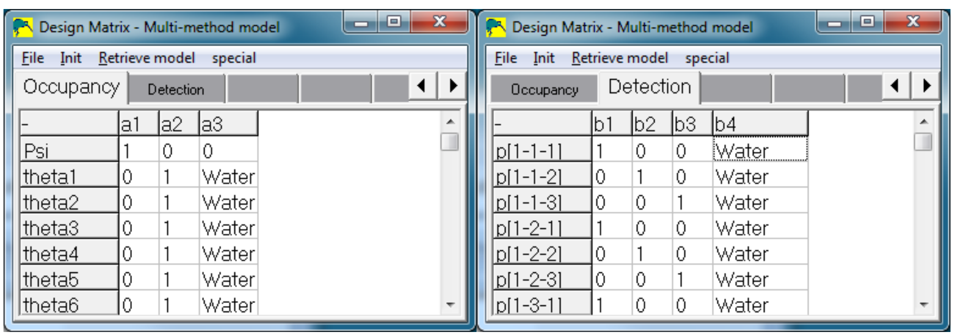
Confusing? OK, for global occupancy, I didn't list an equation for each site because psi doesn't depend on covariates. So, psi is the same for every site. Since local occupancy probability depends on a site covariate, it can be different for each site, so we need to evaluate each equation for every site, plugging in the value of the water covariate for that particular site. The value of the covariate is zero for the first 9 sites, and 1 for the 10^{th} -13^{th} sites. So, I plugged in '0' for sites 1 and 2, and '1' for site 10.

Examine the output for this model, paying attention to the values for local occupancy. Notice that sites 1-9 all have one value for theta, and sites 10-13 have a different value. Does this mean that Nutria are more or less likely to visit a platform on water versus land? What are potential reasons for this? Can you tell this from inspecting the beta coefficients?

Step 6 – Additive effects on detection (p):

Next we'll fit a model where the probability of detection is different for each method as well as whether the platform is on land versus in water, but the effect of land/water is consistent for each method. We'll assume local occupancy is different for platforms on land versus platforms in water and doesn't change from one survey to the next (constant). We could call this model **psi(.)theta(water),p(m+water)**. Again, we'll first fit the model, then work back through it explaining what we're actually doing. So start the analysis by selecting **Run>Analysis:single-season>multi-method**. In the design-matrix window, click Retrieve button on top and select our previous model. This is a good starting point.

Click the 'Detection' tab, then right-click and select 'Add Col'. Click in the last column, then click the '**Init>Indiv.Covariates>*Water**' menu. The design-matrix window should look like this:



Once the design matrices are set up, return to the SNER window (without closing the design matrix window), rename the model (psi(.)theta(water)p(m+water)), then hit 'OK to Run' and confirm the results when prompted.

So what have we just done? Global and local occupancy are estimated exactly as in the previous model, but detection probabilities are method-specific plus an effect of land/water placement. To read the design matrices we move along the row, sum the terms produced by multiplying the values in the grid cell with the corresponding beta parameters for each column. So here we could write the set of equations:

$$\text{logit}(p_{method1,site1}) = \mathbf{1} \text{ X } b_1 + \mathbf{0} \text{ X } b_2 + \mathbf{0} \text{ X } b_3 + \mathbf{Water}_{site1} \text{ X } a_4 = a_1$$

$$\text{logit}(p_{method1,site2}) = \mathbf{1} \text{ X } b_1 + \mathbf{0} \text{ X } b_2 + \mathbf{0} \text{ X } b_3 + \mathbf{Water}_{site2} \text{ X } a_4 = a_1$$

$$\text{logit}(p_{method1,site10}) = \mathbf{1} \text{ X } b_1 + \mathbf{0} \text{ X } b_2 + \mathbf{0} \text{ X } b_3 + \mathbf{Water}_{site10} \text{ X } a_4 = a_1 + a_4$$

: : :

$\text{logit}(p_{method2,site1}) = \mathbf{0} \text{ X } b_1 + \mathbf{1} \text{ X } b_2 + \mathbf{0} \text{ X } b_3 + \mathbf{Water}_{\text{site1}} \text{ X } b_4 = b_2$

$\text{logit}(p_{method2,site2}) = \mathbf{0} \text{ X } b_1 + \mathbf{1} \text{ X } b_2 + \mathbf{0} \text{ X } b_3 + \mathbf{Water}_{\text{site2}} \text{ X } b_4 = b_2$

$\text{logit}(p_{method2,site10}) = \mathbf{0} \text{ X } b_1 + \mathbf{1} \text{ X } b_2 + \mathbf{0} \text{ X } b_3 + \mathbf{Water}_{\text{site10}} \text{ X } b_4 = b_2 + b_4$

: : :

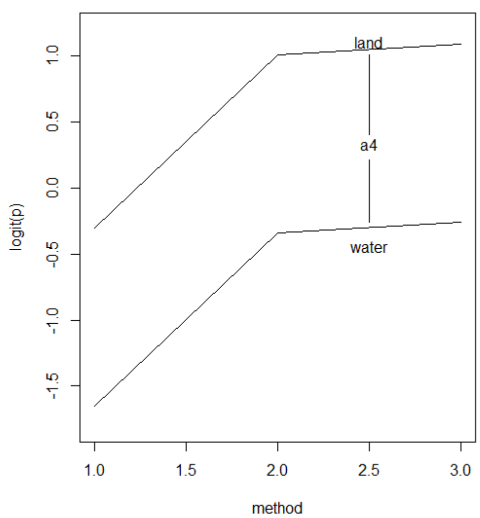
$\text{logit}(p_{method3,site1}) = \mathbf{0} \text{ X } a_1 + \mathbf{0} \text{ X } a_2 + \mathbf{1} \text{ X } b_3 + \mathbf{Water}_{\text{site1}} \text{ X } b_4 = b_3$

$\text{logit}(p_{method3,site2}) = \mathbf{0} \text{ X } a_1 + \mathbf{0} \text{ X } a_2 + \mathbf{1} \text{ X } b_3 + \mathbf{Water}_{\text{site2}} \text{ X } b_4 = b_3$

$\text{logit}(p_{method3,site10}) = \mathbf{0} \text{ X } a_1 + \mathbf{0} \text{ X } a_2 + \mathbf{1} \text{ X } b_3 + \mathbf{Water}_{\text{site10}} \text{ X } b_4 = b_3 + b_4$

: : :

Examining the equations we can see that for platforms on land (Water=0), detection probability is different for each method (a1 for method 1, a2 for method 2, a3 for method 3). Detection probabilities for platforms on water (Water=1) (eg., site 10) are incrementally lower/higher than those on land, where the difference between detection on water versus land is a4. A graph might help:



Looking at the output from this model, a4 was estimated as -1.4 which, as it is <0, indicates that the estimated probability of detection is lower at sites in water versus sites on land. In this fairly simple model we can interpret a4 in a fairly straightforward manner, although for more complicated models it can be useful to interpret the effect of a covariate in terms of an odds ratio. Using the logit link, the odds ratio can be calculated as:

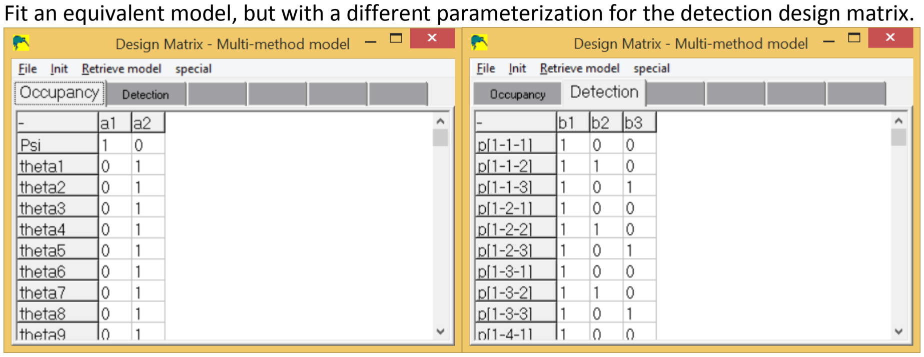
$$OR = e^{a4} = e^{-1.4} = 0.25$$

The interpretation of this would be that the odds of detection are 0.25 times as large for a water site versus a land site. An approximate 2-sided 95% confidence interval for the odds ratio would be:

$$CI = (e^{-1.4 - 2 \times 0.81} , e^{-1.4 + 2 \times 0.81}) = (e^{-3.02} , e^{0.22}) = (0.05, 1.25)$$

Exercise:

1. Working in small groups, fit the model with the following design matrices. Write out the corresponding equations, draw the graphs and interpret the resulting beta parameter estimates. Fit an equivalent model, but with a different parameterization for the detection design matrix.



-	a1	a2
Psi	1	0
theta1	0	1
theta2	0	1
theta3	0	1
theta4	0	1
theta5	0	1
theta6	0	1
theta7	0	1
theta8	0	1
theta9	0	1

-	b1	b2	b3
p[1-1-1]	1	0	0
p[1-1-2]	1	1	0
p[1-1-3]	1	0	1
p[1-2-1]	1	0	0
p[1-2-2]	1	1	0
p[1-2-3]	1	0	1
p[1-3-1]	1	0	0
p[1-3-2]	1	1	0
p[1-3-3]	1	0	1
p[1-4-1]	1	0	0

2. Within your groups complete the following set of models to the nutria data, and discuss the results. Note that some of these you have already completed.
   i. psi(.)theta(.)p(.)
   ii. psi(.)theta(.)p(m)
   iii. psi(.)theta(.)p(water)
   iv. psi(.)theta(.)p(m+water)
   v. psi(.)theta(water)p(.)
   vi. psi(.)theta(water)p(m)
   vii. psi(.)theta(water)p(water)
   viii. psi(.)theta(water)p(m+water).

3. Calculate the summed AIC weights for the *water* covariate for local occupancy(theta) (**hint**: if you right mouse click on the **Results Browser** you can copy the table to the clipboard, and then paste into a spreadsheet).

4. What inferences do you draw about (1) the effect of water on local occupancy and (2) the detection probabilities?

5. Calculate the model averaged estimate (and SE) of the probability of occupancy for a water site. To compute the model averaged estimate, calculate the product of the probability of local occupancy for a water site for each model and the weight of each model. The sum of those products is the model-averaged estimate. The SE is done via the following formula:

Mod-avg-SE(theta) = sum{ w(i) * sqrt((SE(theta(i))^2) + (theta(i)-modavgtheta)^2)) }

Comment [KJ1]: What should they do for models where theta is constant?

Comment [KJ2]: I added in parentheses here as I believe they should be based on having this not work last year. But please double check Jim.

Single Season, Multi-Method Occupancy Case Study – Estimating Occupancy of Nutria in Maryland.